

# Inteligência Artificial na Educação Musical

Anderson dos Santos de Barros<sup>1</sup>, Heverton Pires da Luz<sup>2</sup>, Joas Weslei Baia<sup>3</sup>

## Resumo

Nesse artigo é apresentado como a tecnologia de redes neurais artificiais pode auxiliar o processo de ensino-aprendizagem da música e como realizar o treinamento de uma rede neural para reconhecer notas musicais. O aplicativo foi desenvolvido através da tecnologia Apache Cordova que utiliza os recursos Web: JavaScript, HTML5 e CSS3. Também foi utilizado um servidor baseado na estrutura RestFul, a linguagem de programação Java, os componentes jMusic, NeuroPH e os algoritmos de Transformada Rápida de Fourier (FFT) e Backpropagation. O projeto obteve 94,44% de acerto no reconhecimento das notas musicais treinadas. Além disso, esse projeto disponibiliza um modelo de aplicação de redes neurais artificiais no ensino da música.

**Palavras-chave:** Ensino da música. Redes neurais artificiais.

---

<sup>1, 2, 3</sup>UNIFAGOC - Centro Universitário Governador Ozanam Coelho.

<sup>1</sup>E-mail: anderson.barros0298@gmail.com. Orcid: <https://orcid.org/0000-0002-7497-0195>

<sup>2</sup>E-mail: h.pires03@gmail.com. Orcid: <https://orcid.org/0000-0002-1213-9793>

<sup>3</sup>E-mail: asweslei@gmail.com. Orcid: <https://orcid.org/0000-0002-1050-4750>

## 1 Introdução

A música é uma linguagem de comunicação universal capaz de expressar e registrar a cultura de uma civilização. Ela é uma forma de arte constituída por vários sons e ritmos, seguindo uma organização em relação ao tempo (TEBEROSKY e COLL, 2000).

A computação musical é uma área da Ciência da Computação que estuda a aplicação do computador a problemas musicais. Em 1957, Max Mathews, desenvolveu um software de música, chamado de Music I. Ele controlava afinação, volume e duração dos sons e deu origem a outras três versões, chamadas de Music II, III e IV (FRITSCH et al., 2004).

Esse trabalho mostra como a inteligência artificial pode ser empregada como recurso pedagógico no ensino da música através das seguintes contribuições:

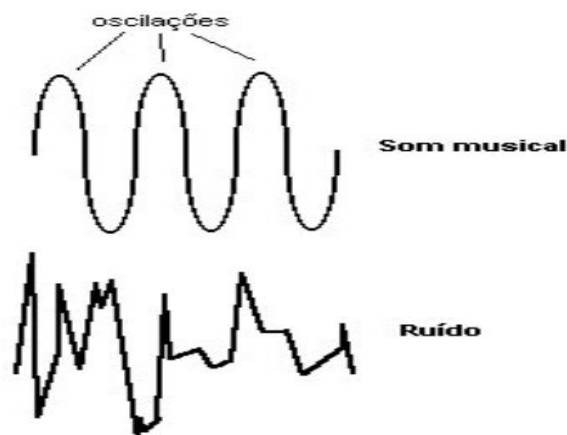
- Como a tecnologia de redes neurais artificiais pode auxiliar o processo de ensino-aprendizagem.
- Como realizar o treinamento de uma rede neural para reconhecer notas musicais.
- Os resultados alcançados no desenvolvimento do projeto.

Esse projeto objetivou auxiliar o aprendizado da música através da identificação de notas musicais usando Redes Neurais Artificiais. O problema resolvido nesse projeto foi o feedback ao aluno sobre o seu desempenho. Por exemplo, esse aluno executando um violão, ele tenta produzir uma nota, o aplicativo realiza a correção indicando o desempenho alcançado. Sua importância está ligada à aprimoração da técnica musical do discente, para que ele possa tocar corretamente as notas musicais, auxiliando-o no processo de aprendizagem.

Esse trabalho tem como objetivo auxiliar os discentes de música através da indicação de sua performance.

## 2 O Som

Segundo (MENEZES et al., 2007), o som é a perturbação do ar como resultado da vibração de um corpo elástico. Ele ainda afirma que as notas musicais têm frequências regulares, que cada uma tem uma frequência específica e que toda a teoria musical no ocidente foi baseada nisso. A Figura 1 mostra a diferença entre uma onda sonora utilizada na música e um ruído.



**Figura 1:** Ondas sonoras  
**Fonte:** MENEZES et al. (2007).

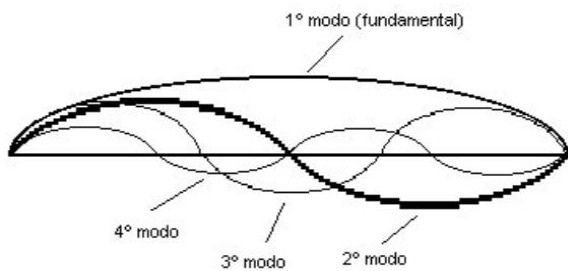
Segundo Farias et al. (2009), os sons musicais têm como características quatro parâmetros:

- Intensidade: representa a potência acústica, mais conhecida como volume.
- Altura: indica o quanto um som pode ser agudo (frequências altas) ou grave (frequências baixas).
- Timbre: também chamado de qualidade, o timbre é uma das características que permite a diferenciação de instrumentos distintos.
- Envelope: também é responsável por distinguir dois instrumentos diferentes. O envelope sonoro é formado por 4 partes distintas: a primeira é o ataque, que é o início de cada nota musical e dependendo do instrumento pode durar alguns centésimos ou até mais de um segundo. A segunda é o decaimento, que acontece em alguns instrumentos,

uma caída de intensidade até um ponto de estabilidade. A terceira é a sustentação, que é tempo de duração da nota musical que muitas vezes pode ser controlada pelo executante, dependendo do instrumento. A última parte é o relaxamento, que é quando o som desaparece completamente, ou seja, o final da nota (LENGNING, 2009).

Com base nesses parâmetros é possível realizar o reconhecimento dos instrumentos musicais e as suas notas.

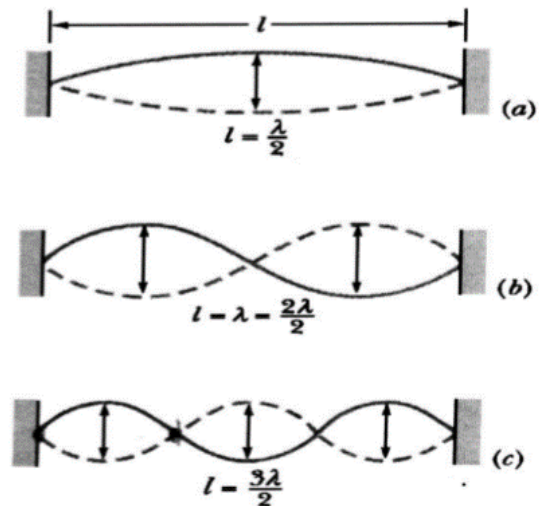
A música é constituída de diversas notas, que são resultantes de vibrações secundárias, essas vibrações definem a formação dos harmônicos. A Figura 2 mostra essas notas ou frequências chamadas de harmônicos (MENEZES et al., 2007).



**Figura 2:** Harmônicos de um corpo elástico  
**Fonte:** MENEZES et al. (2007).

O conjunto de harmônicos que acompanha uma nota musical forma a série harmônica. A Figura 3 mostra a série harmônica gerada ao tocar a nota dó de uma guitarra, a corda vibrará inteira, depois vibrará se dividindo pela metade, depois em três partes iguais e assim por diante (FARIAS et al., 2009).

A nota fundamental que compõe a série harmônica, define qual nota que está sendo tocada. As diferentes intensidades dos harmônicos permitem o reconhecimento dos diferentes timbres.



**Figura 3:** Harmônicos de uma corda  
**Fonte:** FARIAS et al. (2009).

### 3 Representação Digital do Som

Segundo (FRITSCH et al., 2004) a digitalização é o processo de representar numericamente grandezas analógicas (áudio, vídeo e imagem) em computadores, que são máquinas que trabalham na notação binária. Com isso foram criados dois circuitos para tornar possível a conversão de um sinal analógico para digital e a realização do processo inverso, a conversão do sinal digital em analógico. Os circuitos foram chamados de circuitos analógico-digital (*Analogic-Digital Converter - ADC*) e digital-analógico (*Digital-Analogic Converter - DAC*). Segundo (MORETTI, 2003), o processo de digitalização do som, é realizado em quatro etapas:

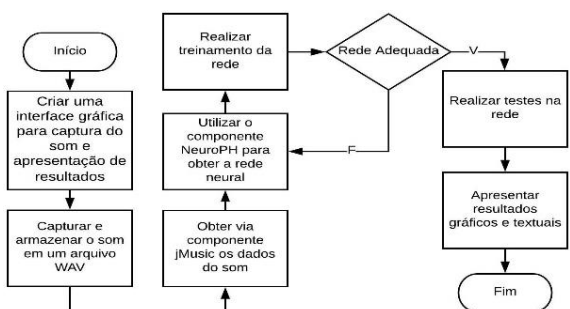
- Amostragem: processo que determina um número de pontos de amplitude da onda e a conversão desses pontos para números que representem a amplitude.
- Filtragem: limita uma faixa de frequência, onde valores acima da frequência máxima são descartados.
- Quantização: arredonda os valores de nível de amplitude para o valor mais próximo quando ficam entre dois valores digitais, ou seja, para um computador que trabalha com dois dígitos, os valores de amplitude maior

ou igual a 10,5 e menor que 11,5, serão arredondados para 11.

- Por fim, realizar a gravação dos dados obtidos em arquivo.

#### 4 Método de Desenvolvimento do Trabalho

Foram desenvolvidas sete atividades para atingir os objetivos desse trabalho, conforme ilustra a Figura 4.



**Figura 4:** Método de desenvolvimento do trabalho

1. As interfaces gráficas com o usuário foram criadas utilizando a tecnologia híbrida, com a plataforma Apache Cordova que utiliza HTML5, CSS3 e Java Script no desenvolvimento de aplicativos móveis para permitir que o aplicativo seja usado em dispositivos Android e IOS. Os resultados dos cálculos são apresentados ao usuário de forma gráfica e textual. Primeiramente é apresentada a nota que foi reconhecida, juntamente com gráfico de percentual, que mostra o percentual de acerto, ou seja, a porcentagem de chance daquela nota reconhecida ser realmente a nota executada.
2. Para a captura da nota musical o usuário utiliza a interface gráfica que permite o acionamento do microfone do dispositivo e a respectiva gravação. O áudio coletado é armazenado em um arquivo WAV que é a entrada do Webservice que foi desenvolvido utilizando a

linguagem Java. Ele utiliza dos componentes jMusic, NeuroPH e AudioAnalsi, para realizar o reconhecimento das notas musicais e retornar os resultados para o aplicativo.

3. Para o processamento dos dados relativos ao som capturado, foi utilizado o componente Java jMusic, pois ele oferece recursos e funções necessários para o projeto: a leitura do arquivo WAV, a taxa de amostragem, a quantidade de canais e os valores que representam as notas da música.
4. Foi concebida uma Rede Neural Artificial para aprender as notas musicais. Nessa etapa, a rede foi configurada da seguinte forma: na camada de entrada foram colocados 44097 neurônios, na camada oculta 8 neurônios e na camada de saída foram 6 neurônios, conforme apresentado na Figura 6.
5. A partir dos dados numéricos que representam o som de uma nota musical, foi realizada a etapa de treinamento da rede neural. Esses dados são acessados pelo jMusic e repassados para o algoritmo de Transformada Rápida de Fourier (FFT) que é fornecido pelo componente AudioAnalsi. Esse algoritmo é necessário tendo em vista que os dados devem representar o domínio da frequência do som emitido. O resultado é repassado para rede neural para que ela aprenda o padrão daquela nota, e posteriormente reconhecê-la em outro conjunto de dados, ou seja, quando aquela nota for produzida pelo instrumento a rede irá reconhecê-la.
6. Foram capturadas amostras das notas musicais de um violão para a realização dos testes. Foi reproduzido o som da nota do violão em um computador e esse som foi gravado com o aplicativo, para que

ele reconhecesse a nota, simulando uma atividade real de uso do aplicativo.

- Os resultados foram calculados utilizando as Equações 1, 2 e 3:

$$Erro = \sqrt{\sum_{i=1}^6 (resultadoEsperado[i] - resultadoObtido[i])^2}$$

**Equação 1:** Erro Quadrático

A Equação 1 realiza o cálculo do quadrado da diferença entre o resultado esperado, que representa a nota musical em seis dígitos, e o resultado obtido, representado pela saída da Rede Neural Treinada ao apresentar o som de uma dada nota musical.

$$PE = \frac{(Erro * 100)}{2,45}$$

**Equação 2:** Erro Percentual

$$PA = (100 - PE)$$

**Equação 3:** Porcentagem de Acerto

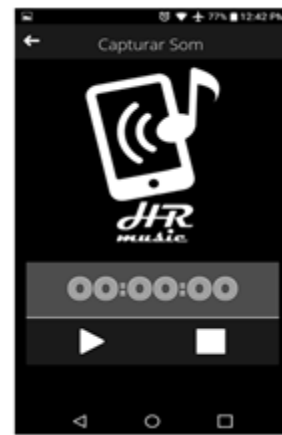
A Equação 2 utiliza a divisão por 2,45, é o valor do erro máximo possível, ou seja, 100% de erro no reconhecimento. Ele foi calculado com base na Equação 1, ou seja, foi encontrada a diferença máxima possível entre a representação de uma nota e seu valor calculado. Essa avaliação ocorre a cada 1/5 do segundo, ou seja, a cada 0,2 segundos.

**5 Resultados e Discussão**

O aplicativo desenvolvido nesse trabalho, denominado HRmusic, foi implementado com o Visual Studio 2017 e a plataforma Apache Cordova, esta última utiliza as linguagens Java Script, CSS3 e HTML5, necessárias para a utilização de funções e periféricos nativos de um dispositivo móvel: câmera, microfone e sensores. Uma vantagem da utilização desse tipo de tecnologia é que o mesmo código-fonte pode ser utilizado em várias plataformas: Android, Windows Phone e IOS.

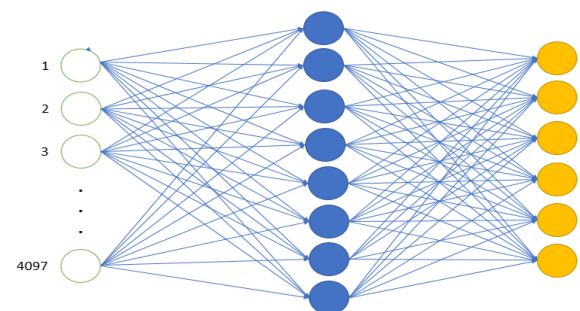
A Figura 5 mostra a interface gráfica para a captura dos sons emitidos pelo aluno. Ao

pressionar o botão de início da gravação, o sistema dispara uma mensagem que bloqueia a tela por 10 segundos, com intuito de possibilitar ao usuário um tempo para posicionamento do instrumento. A mensagem pode ser ignorada com um simples toque na tela e assim imediatamente dar início a gravação do som, ou pelo contrário a gravação só se inicia ao passar dos 10 segundos.



**Figura 5:** Interface para captura do som do instrumento musical

A rede neural artificial foi concebida com 4097 neurônios na camada de entrada, 8 neurônios na camada oculta e 6 neurônios na camada de saída, conforme mostra a Figura 6.



**Figura 6:** Rede Neural Artificial desenvolvida

Esta Rede Neural Artificial tinha inicialmente uma configuração de 44100 neurônios na camada de entrada, 4410 neurônios na camada oculta e 6 neurônios na camada de saída. Porém, o tempo de treinamento tornou-se inviável com o hardware disponível. A partir desse problema, a rede foi ajustada na quantidade

de neurônios até chegar na configuração da Figura 6. Os dados do áudio foram submetidos ao algoritmo de FFT, assim a rede sofreu um novo ajuste na camada de entrada, pois o algoritmo por definição só aceita como entrada um vetor que tenha um tamanho que seja múltiplo de 2, ou seja, o número de amostras do som emitido foram configuradas para 8.192 para que pudesse ser feita a entrada de dados na FFT e o retorno do algoritmo ter um tamanho que é a metade do vetor que foi passado na entrada (4097 amostras). Com isso, a rede chegou na configuração ideal com 4097 neurônios na camada de entrada, passando a receber como entrada o resultado do algoritmo da FFT.

Para realizar o processamento e o reconhecimento das notas foi desenvolvido um Webservice que responde a chamadas HTTP, assim processamentos que exigem mais recursos da máquina, como treinamento da rede, manipulação de arquivos de áudio, cálculos estatísticos, reconhecimento das notas podem ficar em um servidor e o aplicativo fica apenas com processamentos mais leves como: gravar e enviar os arquivos de som e apresentar os resultados para o usuário. A integração entre o Webservice e o aplicativo é dado pela estrutura da Figura 7.



**Figura 7:** Estilo arquitetônico da aplicação

Quando é solicitada uma gravação do áudio de uma nota musical, a aplicação realiza uma requisição GET ao servidor, que caso esteja online, ele inicia os parâmetros da rede neural. Após a conexão ser estabelecida e as configurações serem carregadas, o Webservice está pronto para receber a requisição POST, com o arquivo de som.

No aplicativo, quando o usuário para a gravação, automaticamente o sistema salva o arquivo no celular e realiza a requisição ao servidor passando o arquivo que foi gerado, para que ele possa ser processado e reconhecido pela rede neural. Após terminar

o processamento, o servidor retorna o resultado do processamento para a aplicação.

## 6 Aprendizagem

Foram treinadas e testadas doze notas musicais, resultando no reconhecimento de dez notas. Os resultados representados na Tabela 1, são referentes aos primeiros 0.2 segundos de reconhecimento, sendo que as notas FA e FÁ# são as duas notas que não foram reconhecidas em nenhum instante. Entretanto, as notas RÉ e SOL foram reconhecidas em alguns intervalos diferentes de 0.2 segundos.

<u>Toc</u>	<u>Rec</u>	<u>Erro (%)</u>	<u>Acert (%)</u>	<u>Temp (s)</u>
DO	DO	5,56	94,44	0,2
DO#	DO#	7,75	92,25	0,2
FA	RE#	35,46	64,54	0,2
FA#	DO	5,56	94,44	0,2
LA	LA	21,53	78,47	0,2
LA#	LA#	6,14	93,86	0,2
MI	MI	15,12	84,88	0,2
RE	MI	14,92	85,08	0,2
RE#	RE#	5,89	94,11	0,2
SI	SI	4,48	95,52	0,2
SOL	DO#	4,64	95,36	0,2
SOL#	SOL#	19,71	80,29	0,2

**Tabela 1:** Resultado do aprendizado artificial

### Legenda:

**Toc.** – Nota tocada.      **Rec.** – Nota Reconhecida

**Acert.** – Percentual de acerto.

**Erro** – Percentual de erro no reconhecimento da nota.

**Temp.** – Instante de reconhecimento da nota em segundos.

Quando a nota FÁ foi apresentada à Rede Neural Artificial treinada, foi reconhecida a

nota RE#, em 0.2 segundo de reconhecimento. Quando a nota foi produzida, houve a captura de um ruído.

Na Tabela 1 é apresentado a probabilidade de acerto, que está relacionada com o resultado esperado e o resultado obtido. Apesar da nota reconhecida estar errada, o sistema apresenta uma probabilidade de acerto de 64,54%. Isso acontece, pois, no momento que inicia o reconhecimento da nota o sistema não sabe qual é a nota que está sendo tocada. Então ele realiza uma busca em todas as notas aprendidas e calcula a taxa de erro de cada uma, levando em conta o esperado e o resultado da execução do processamento, assim a cada execução, é verificado se o erro calculado é menor que o erro armazenado anteriormente e caso isso seja verdade, o sistema realiza a atualização do erro, obtendo no final o menor erro dentre todas as notas. O cálculo do erro foi realizado utilizando a Equação 1, o quadrado da diferença entre o valor esperado e o valor obtido.

A avaliação do reconhecimento do intervalo de gravação de 2 segundos de áudio da nota musical DÓ é apresentado na Tabela 2. Durante esse intervalo de tempo a rede neural artificial obteve uma média de 89,46% de acerto.

Nos instantes de 1.4 e 1.6 segundos, a porcentagem de acertos diminuiu para 65,24% e 73,92%, devido aos ruídos captados nesses instantes de tempo.

Analisando a Tabela 1, resultado dos primeiros 0.2 segundos, observa-se que a nota RÉ foi reconhecida como a nota MI. Porém, após esse intervalo, a nota foi reconhecida com uma taxa média de acerto de 94,39%. A Tabela 3 permite a análise desse resultado.

A avaliação do reconhecimento do intervalo de gravação de 2 segundos de áudio da nota musical DÓ é apresentada na Tabela 2. Durante esse intervalo de tempo a rede neural artificial obteve uma média de 89,46% de acerto.

N	Acert (%)	Temp (s)
	94,44	0,2
	94,44	0,4
	94,44	0,6
	94,44	0,8
DÓ	94,44	1,0
	94,44	1,2
	65,24	1,4
	73,92	1,6
	94,44	1,8
	94,44	2,0

**Tabela 2:** Resultado da execução da nota DÓ no intervalo de 2 segundos.

N	Acert (%)	Temp (s)
MI	85,08	0,2
	85,08	0,4
	92,18	0,6
RE	96,61	0,8
	96,61	1
	92,18	1,2
MI	89,50	1,4
	96,61	1,6
RE	96,61	1,8
	92,18	2

**Tabela 3:** Resultado de reconhecimento das notas RÉ e MI no intervalo de 2 segundos.

Os dados dos resultados calculados e apresentados de forma tabular são

ilustrados de forma gráfica na interface gráfica do aplicativo, conforme mostra a Figura 8, que apresenta o resultado de quatro testes. A primeira informação mostra a nota reconhecida pela rede neural artificial no intervalo de 0.2 segundo.



**Figura 8:** Apresentação do resultado computado

A segunda informação mostra a probabilidade de acerto, ou seja, a chance de a nota tocada ser a nota reconhecida. Por último, o gráfico de linhas, comparando o resultado obtido com o resultado esperado, representados pelas cores verde e vermelho, respectivamente. Os botões presentes no rodapé da tela têm a função de navegar entre os vários 0.2 segundos que um áudio pode gerar.

A Figura 8 exhibe o resultado do reconhecimento das notas musicais SI, DO,

RE# e DO#, apresentado após o término do upload do arquivo para servidor, onde é realizado o processamento e então é retornado o resultado do reconhecimento para a exibição na tela do aplicativo.

## 7 Modelo de Aplicação de Redes Neurais Artificiais no Ensino da Música

Foi concebido um modelo para a aplicação da tecnologia de Redes Neurais Artificiais no Ensino da Música compostos pelas seguintes atividades:

- Concepção do exercício para o aluno usando a pauta musical, por exemplo, o professor disponibiliza o exercício sobre a escala cromática do instrumento. O projeto desse requisito funcional foi concebido mediante o uso do componente de software VexFlow<sup>1,2</sup>.
- Para enviar e receber os exercícios, conforme ilustrado na Figura 7, foi adotada representação através do MusicXML<sup>3</sup>, um modelo de intercambio de partituras digitais.
- O aluno acessa o servidor e obtém a página com o referido exercício. Ele realiza o exercício e recebe o resultado com seu desempenho, semelhante ao ilustrado na Figura 8.
- Conforme o resultado obtido, ele pode analisar quais erros foram introduzidos, e então repetir o exercício com a possível correção. Além disso, o resultado fica hospedado no servidor para acompanhamento futuro.
- O professor pode acompanhar o resultado do aluno através de relatórios que indicam o status de seu progresso. Essa informação pode ser usada para subsidiar o planejamento da aula presencial.

<sup>1</sup> <http://www.vexflow.com/>

<sup>2</sup> <https://github.com/Oxfe/vexflow>

<sup>3</sup> <https://www.musicxml.com/>



## 8 Considerações Finais

A Rede Neural Artificial concebida nesse trabalho obteve um desempenho médio de 94,44% de acerto no reconhecimento das notas musicais, conseguindo reconhecer 10 notas das 12 notas treinadas.

No projeto foram utilizadas as ferramentas Bitbucket e SourceTree para controle e versionamento de código fonte, além do emprego da Engenharia de Software baseada em reuso.

Sugere-se como trabalhos futuros a construção de filtros de ruído para melhorar o reconhecimento e o suporte para o reconhecimento de outras notas e outros instrumentos.

## 9 Referencias

FARIAS, Anderson Jean; LINDEN, Jorge Fernando; TERNUS, Marcio Alessandro; SOUZA, João Olegário O. **Identificação de Instrumentos Musicais Utilizando Redes Neurais Artificiais**. Revista Liberato, Novo Hamburgo, v. 10, n. 13, p.91-101, jan./jun. 2009.

FERNANDES, Luiz Gustavo L.; PORTUGAL, Marcelo S. **Redes Neurais Artificiais e Previsão de Séries Econômicas: Uma Introdução**. Nova Economia, Belo Horizonte, v.6, n. 1, 1996.

FRITSCH, E. F., MILETTO, E. M.; CASTALONGA, L. V. Flores; PIMENTA, M. S.; VICARI, R. M. **Introdução à Computação Musical**. In: CBComp - Congresso Brasileiro de Computação, 4., 2004, Itajaí. **Anais...**, p. 883-902. Itajaí, SC - Brasil, ISSN 1677-2822.

LENGNING, Rodrigo. **O que são envelopes sonoros?** AIMEC: Academia Internacional de Música Eletrônica, Curitiba, dez. 2009. Disponível em: <<https://www.aimec.com.br/o-que-sao-envelopes-sonoros/>>. Acesso em: 23 dez. 2018.

MENEZES JUNIOR, Carlos R.F.; FARIA, Eustáquio José; YAMANAKA, Keiji.

**Reconhecendo instrumentos Musicais Através de Redes Neurais Artificiais**. Hifen, Porto Alegre, v. 31, N. 59, 2007.

MORETTI, Adriano Luiz. **Protótipo de um Software para o Reconhecimento de Notas Musicais**. Trabalho de Conclusão de Curso, Universidade Regional de Blumenau, Blumenau, 2003.

TEBEROSKY, Ana; COLL, César. **Aprendendo Arte**. São Paulo: Ática, 2000.